AD-A067 518 NAVAL RESEARCH LAB WASHINGTON D C F/G 6/4
ARTIFICIAL INTELLIGENCE AND ITS CRITICS.(U)
FEB 79 T L JONES
UNCLASSIFIED NRL-MR-3927 SBIE-AD-E000 288 NL

1 OF 1
ADA
067518

END
DATE
FILMED
6 79
DDC

ADE 000 288

NRL Memorandum Report [illegible]

AD A067518

Artificial Intelligence and Its Critics

THOMAS L. JONES

Computer Science Laboratory
Communications Sciences Division

LEVEL III

DDC FILE COPY

February 23, 1979

NAVAL RESEARCH LABORATORY
Washington, D.C.

[illegible]

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Memorandum Report 3927	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ARTIFICIAL INTELLIGENCE AND ITS CRITICS		5. TYPE OF REPORT & PERIOD COVERED Interim report on a continuing NRL problem.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas L. Jones*		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NRL Problem B02-48
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE February 23, 1979
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

*Present address: 1428 Hampshire Court West, Silver Spring, Maryland 20903

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computers
Artificial Intelligence
Limitations of Computers
Computer Programs
Computer Applications

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It has been about 20 years since the first papers in A.I. appeared. In this period, A.I. research has produced many interesting ideas of which list processing is probably the most widely used. Other results of A.I. research include chess programs, etc. Progress has been slower than some people predicted. Therefore, there has developed a body of literature which claims that A.I. effort is misguided. In this paper I will reply to three prominent critics of A.I., Dreyfus, Weizenbaum, and Lighthill.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

79 03 09 021

ARTIFICIAL INTELLIGENCE AND ITS CRITICS

It has been about twenty years since the first papers in Artificial Intelligence appeared; e.g., (12,17,25). In this period, Artificial Intelligence research has produced many interesting ideas, of which list processing is probably the most widely used. Other results of artificial intelligence (A.I.) research include chess programs (16), programs which solve calculus problems (27) and a program which derives a molecular structure from a spectrogram (8). Progress has been slower than some people predicted. In consequence, there has developed a body of literature which claims that the entire A.I. effort is misguided. In this paper, I will reply to three prominent critics of A.I., Dreyfus, Weizenbaum, and Lighthill. Although I find little merit in the attacks of these critics, A.I., like literature or philosophy, is a difficult enough endeavor that it can benefit from intelligent criticism. Accordingly, I will try to provide some.

The first attack on A.I. to receive widespread publicity was by Hubert L. Dreyfus, as described in his book What Computers Can't Do (6). Dreyfus is a philosopher who specializes in existentialism and phenomenology. Although he has read extensively in computer science, Dreyfus does not seem to have the detailed technical expertise which comes from actually performing computer research. He makes a

Note: Manuscript submitted December 14, 1978.

complicated argument which will only be sketched here.

He lists several characteristics of human thought and claims that they cannot be replicated in a "step-by-step" computer. In particular, he claims that computer behavior is "rule-like" in a way that human behavior is not. He illustrates by comparing a horse player who uses a system with an intuitive horse player who goes by guess and the Racing Form (5). Intuitive horse playing is claimed by Dreyfus to be an "UHFIP," or "Uniquely Human Form of Information Processing," to use Papert's phrase (18). It seems to me that Dreyfus falls into a philosophical trap by failing to distinguish between two senses of the notion "behavior conducted according to rules." In an obvious sense, the system player goes by rules (the system), while the intuitive player does not. In a second sense, the behavior of the intuitive player can, in theory, be described in terms of the laws of psychology and neurophysiology, although of course we do not know these laws well enough to predict the intuitive horse player's behavior. A rule of type I is something adopted by people as a guide for conduct, such as an ethical code. A rule of type II is a law of nature which describes the behavior of a (living or non-living) thing. Thus the motion of the planets is rule-like in the sense of type II but not rule-like in the sense of type I. The notion that we humans are, in principle, describable by the laws of nature (rules of type II) is the foundation

of psychology and neurophysiology. In a second argument, Dreyfus compares the brain with an analog computer, suggesting that digital computers are stupid because of their step-by-step character. But a continuous process can be approximated by a digital computer to any desired degree of accuracy, at a considerable price in speed of computation. Dreyfus would be on firmer ground if he attacked A.I. based on philosophical arguments (e.g., that machines lack consciousness and free will). It is ironic that Dreyfus, the philosopher, considers the technical question of whether A.I. <u>can</u> be achieved, while Weizenbaum, the computer scientist, studies the ethical and political question of whether A.I. <u>should</u> be achieved.

If Dreyfus is un-knowledgeable about computers, the opposite is true of Joseph Weizenbaum, author of <u>Computer Power and Human Reason</u> (33). Weizenbaum is a distinguished computer scientist and inventor of the well-known Eliza program (32), which carries on a rudimentary conversation with the user. The crux of Weizenbaum's argument is social and political. He argues that computers have reinforced the "most conservative, indeed, reactionary, ideological components of the (spirit of the age)" (34). Making frequent references to the role of computers in the Vietnam war, he argues that attempts to replace human reason by computers are "obscene". One illustration is the work of

Kenneth Colby (2) in applying computers to psychiatry. Anyone who has visited a large state mental hospital will testify that these institutions are obscene because of under-funding and lack of adequate therapy. If we had a way of getting the computer to carry on the simplest conversation with a withdrawn, institutionalized patient, it would be a blessing. Weizenbaum criticizes this work on the ground that it is obscene to replace the psychiatrist with a computer; after all, the computer, not being human, cannot really empathize with the patient. A more cogent criticism is that the state of the art of language understanding by computer gives the computer hardly a glimmer of understanding about what the patient is saying. Weizenbaum seems to be making a blanket indictment, not only of A.I. research, but of computers in general and society as a whole. Our society has its flaws, but this does not mean that A.I. research is wasted effort.

Sir James Lighthill is a prominent mathematician who is perhaps best known for his work on generalized functions(10). Lighthill has written a critique of A.I. (11) which at least attempts to be fair (more than can be said for Dreyfus or Weizenbaum). Lighthill divides A.I. research into three areas: A, B, and C. A is for advanced automation; C is central nervous system theory and simulation. B is the bridge activity, which Lighthill identifies with building

robots.

Lighthill supports research in advanced automation, but not particularly by A.I. researchers. He also supports research in simulation of the nervous system, but emphasizes the role of psychologists and neurophysiologists. He criticizes robot research as unsound. I agree that it would be desirable if psychologists and neurophysiologists made more use of computers. But behavior modeling needs a very capable programmer with a thorough knowledge of heuristic methods; it places quite a strain on his programming ability. There are not many psychologists and neurophysiologists willing and able to write this type of program.

Despite Lighthill's attack, robotics remains the best way to study computer interaction with the world, a key feature of intelligence. The robotics studies on perception and motor control rank among the best studies of these topics in any branch of science (33). Lighthill doesn't seem to grasp the need for basic research in A.I. which can support advanced automation and brain simulation. Sutherland (29) points out that some of the most interesting innovations which have come from A.I. came from the bridge (or basic) area.

Having criticized the critics, I will now turn to a much more important activity: criticizing A.I. itself. Why has progress been slow? We need not look far for an answer: Progress is slow because we are attacking a very basic and very difficult scientific problem, that of understanding intelligence.

Based on twenty years' experience, what are the prospects for machine intelligence? I shall begin by dividing the set of all problems which one might pose for an information processor into four categories: Hardware problems, ordinary (programmable) computer problems, a category which I will call "learnable problems," and, finally, unsolveable problems. Hardware problems are simple enough to be solved with logic or analog components and need no digital computer. Ordinary computer problems are defined as those for which it is practical to write by hand a program for solving the problem. Learnable problems are those which it is impractical to program by hand and, with current technology, must be solved by people. Unsolveable problems can be solved by neither man nor machine; an example is predicting, with absolute generality, whether or not a computer program will halt; see Minsky (13).

The class of all learnable problems may be subdivided into two categories: The problems which a computer can

learn to solve, using a few thousand dollars worth of learning code, and those which, in principal, require a <u>person</u> to solve them; I will call these the "person-requiring" problems. Examples of such are:

(1) Fully automatic, high-quality translation of natural languages.

(2) Equalling human ability to learn from experience.

(3) Equalling human ability to remove ambiguities from visual and speech inputs.

Readers who are familiar with automata theory may object to the idea of a "person-requiring" problem on the ground of Turing's thesis (14), which asserts that any "well-defined" procedure can be carried out, in principle, by a class of machines called Turing machines, of which the digital computer is an example. But I shall argue that the "person-requiring" problems cannot be solved by machines without turning the machine into a person, which is possible in principle if Turing's thesis is correct. However, to turn the machine into a person is far beyond the state of the art, to the point of being science fiction.

Let us begin by comparing people and machines, viewed as "information processors". It is common sense that computers are fairly stupid and that people are smarter. The reasons are far from obvious, indeed far from being understood. A.I. researchers have come to realize that one key to good problem solving is to have good knowledge about the problem and the methods for solving it. Part of the

reason why humans are smarter than computers is simply that we know more.

(1) People are apparently born "knowing" quite a lot about how to deal with the world. (See Munn (15) for a summary of research in this area.) Such abilities as walking, grasping, and extracting grammar rules (1) seem to be supported by a repertoire of innate abilities. See, for example, Dennis (4) on "learning" to walk.

(2) People have been to "playpen university." "Playpen university" refers to the education a child receives in the first few years of his life, concerning space, time, objects, causality, and people; see Piaget (19,20,21). The computer, lacking an education from "playpen university", is forced to rely on a laboriously coded program for information about what interpretations of the world "make sense."

(3) People are socialized into the values, language, and knowledge of a particular culture. A young person in our culture goes through years of schooling which the computer notably lacks. Some experts, such as Vygotsky (30), believe that natural languages provide key concepts used in describing and solving problems.

Keeping in mind the idea that only some problems are programmable by hand, let us look at some areas of A.I. research and try to assess what can be expected from them in the near future. Computer Vision is one of the

most intensely studied subfields in A.I. A basic difficulty in competing with humans in this area is that humans can draw on extensive experience about what things look like. There have been numerous attempts to get computers to learn to recognize objects (7). These programs "see" less well than the programs which do not learn at all. This is apparently because, to get the computer to learn to recognize a pattern (i.e., to automatically code a recognition program for the pattern), one must first know the organization and format of the recognition algorithm. Thus programs which see a limited number of objects quite well are a necessary preliminary to machines which learn to see.

Computer robotics suffers from the above-mentioned problems with computer vision. In addition, the code for getting a robot to solve a problem tends to grow, cancerlike, into many thousands of words. For example, Shaky (22), the robot vehicle developed at Stanford Research Institute, needed 250,000 words of code and data (23) to push a block off a platform. There is still hope of practical results in robotics. Perhaps we will end the spectacle of people spending grim days on the assembly line putting automobiles together (probably a programmable problem).

Computer chess is of interest as a way to study planning and look-ahead. The later chess programs are

fairly "brute-force," inspecting several thousand positions in deciding on a move. By comparison, a human chess master looks at 100 or fewer positions. Simon and Chase (26) suggest that a human master can recognize and deal appropriataly with perhaps 50,000 different types of positions. Current hardware is probably too serial and hence too slow to search 50,000 pattern, even if we could assemble the file of position types. There are algorithms such as hash coding, which can search 50,000 positions quickly. However, these algorithms color the search process in ways which may be undesirable.

Future historians may consider that the importance of A.I. research is in its contributions to our understanding of human though and behavior, more than in its practical, technological impact. In my opinion, the best research on problem-solving, planning, vision, and learning since 1955 has come from this little-publicized, much-criticized branch of computer science. There are signs of a rapprochement between A.I. and cognitive psychology. A.I. work in language understanding, in particular, seems to be influencing workers in more traditional fields.

A basic difficulty with A.I. is that we must still <u>program</u> the computer, telling it in laborious detail just what to do. The cost and difficulty of programming are serious obstacles to practical applications of A.I. There is only the most primitive technology of getting the

computer to learn by experience how to do interesting things (9,28,31). Machine learning is an old idea which never got very far. The early approach was to get machines to learn chess playing or other adult behavior. The approach failed, presumably because the computer, not having been to "playpen university," lacked the concepts and methods for learning adult skills.

Perhaps the best approach to machine learning is to try to program the machine to learn perceptual and motor skills, such as recognizing objects and moving them around. Winston (35) has written a program which can learn to recognize simple objects, such as arches, by generalizing from correctly labeled examples. The pitfall to avoid in studying machine learning is the "bootstrap from nothing" syndrome in which the researcher tries to give the computer no information at all, with the idea that the machine can learn everything for itself. Reference (9) describes an approach to machine learning. The idea is to have a repertoire of primitive operations, such as recognizing when one event causes another, plus learning code that senses when the primitive operations can profitably be put into learned program. If this approach proves feasible, it would provide an alternative to programming the machine by hand to solve every detail of a complex problem.

In summary, I believe that what A.I. needs is a truly

automatic way of generating a computer program. When the day comes when we can <u>train</u> the machine, rather than program it, then A.I. will become a practical tool rather than a research effort.

Acknowledgements: I would like to thank Dr. John K. Dixon and Dr. Rodney Johnson for their critical readings of the manuscript.

References

(1) Chomsky, A.N., Aspects of the Theory of Syntax, MIT Press, Cambridge, Mass., 1965.

(2) Colby, K.M., J.B. Watt, and J.P. Gilbert, "A Computer Method of Psychotherapy: Preliminary Communication," J. Nervous and Mental Disorders, Vol 192, No. 2, pp. 148-152; 1966.

(3) Darrach, Brad, "Meet Shaky, the First Electronic Person," Life, Vol. 66, p.58; Nov. 20, 1970.

(4) Dennis, W., "The Effect of Cradling Practices on the Onset of Walking in Hopi Children," J. Genetic Psychology, Vol. 56, pp 77-86; 1940.

(5) Dreyfus, H.L., "Alchemy and Artificial Intelligence," Rand Corporation Paper P-3244; December, 1965.

(6) Dreyfus, H.L., What Computers Can't Do, Harper and Row; 1972.

(7) Duda, R.O. and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, N.Y.; 1973.

(8) Feigenbaum,E., "Artificial Intelligence: Themes in the Second Decade," Proc. IFIP Congress, Vol. II, pp. 1000-1008, Spartan Books, Washington, D.C.; 1968.

(9) Jones, T.L., "A Computer Model of Simple Forms of Learning in Infants," Proc. AFIPS 1972 Spring Joint Computer Conference;p. 885.

(10) Lighthill, M.J., "Introduction to Fourier Analysis and Generalized Functions," University Press, Cambridge, England; 1958.

(11) Lighthill, M.J., "Artificial Intelligence: A General Survey," in Artificial Intelligence: A Paper Symposium, British Science Research Council; 1973.

(12) Minsky, M.L., "Heuristic Aspects of the Artificial Intelligence Problem," Group Report 34-55, Lincoln Laboratories, MIT; December, 1956.

(13) Minsky, M.L., Computation: Finite and Infinite Machines, p.148. Prentice-Hall, Englewood Cliffs, N.J.; 1967.

(14) Minsky, op. cit., p. 108.

(15) Munn, N.L., The Evolution and Growth of Human Behavior, Houghton Mifflin Co., Boston, Mass.; 1955.

16) Newborn, M., Computer Chess, Academic Press, New York, N.Y.; 1975.

(17) Newell, A.R., and H.A. Simon, "The Logic Theory Machine: A Complex Information Processing System," IRE Trans. on Information Theory, Vol. IT-2, No. 3, pp.61-79; 1956.

(18) Papert, S., "The Artificial Intelligence of Hubert L. Dreyfus: A Budget of Fallacies," MIT Artificial Intelligence Memo No. 154; January, 1968.

(19) Piaget, J., The Construction of Reality in the Child, Basic Books, New York, N.Y.; 1954.

(20) Piaget, J., Play, Dreams, and Imitation in Childhood, Norton, New York, N.Y.; 1962.

(21) Piaget, J., The Origins of Intelligence in Children, International Universities Press, New York, N.Y.;1952.

(22) Raphael, B., "Programming a Robot," Proc. IFIP 1968 Congress, Edinburgh, Scotland; August, 1968.

(23) Raphael, B., personal communication; February, 1977.

(24) Reddy, D.R., L. Erman, and R. Neely, "A Mechanistic Model of Speech Perception," Proc. 1972 Conf. on Speech Processing; IEEE, pp. 334-337; 1972.

(25) Shannon, C.E., "Programming a Computer for Playing Chess", Philosophical Magazine, Vol. 41, pp.256-275; 1950.

(26) Simon, H.A., and W.G. Chase, "Skill in Chess," American Scientist, Vol. 61, No. 4, pp.394-403;July-August, 1973.

(27) Slagle, J.R., "A Heuristic Program that Solves Integration Problems in Freshman Calculus," in Computers and Thought, Feigenbaum and Feldman, Eds., McGraw-Hill Book Co., New York, N.Y.; 1963.

(28) Sussman, G.J., A Computer Model of Skill Acquisition, American Elsevier Publishing Co.; 1975.

(29) Sutherland, N.S., "Some Comments on the Lighthill Report and on Artificial Intelligence", in Artificial Intelligence: A Paper Symposium, British Science Research Council; April, 1973.

(30) Vygotsky, L.S., Thought and Language, MIT Press, Cambridge, Mass.; 1962. Originally published in 1934.

(31) Waterman, D.A., "Generalization Learning Techniques for Automating the Learning of Heuristics," Artificial Intelligence, Vol. 1, Nos. 1 and 2; Spring, 1970.

(32) Weizenbaum, J., "ELIZA--A Computer Program for the Study of Natural Language Communication Between Man and Machine," Comm. ACM, Vol. 9, No. 1, pp. 36-45; January, 1965.

(33) Weizenbaum, J., Computer Power and Human Reason, W.H. Freeman and Co., San Francisco, Calif.; 1976.

(34) Weizenbaum, J., op. cit., p. 227.

(35) Winston, P.H., "The MIT Robot," in Machine Intelligence 7, Meltzer and Michie, Eds., John Wiley and Sons, New York, N.Y.; 1972.